As Time Goes By:
How Period Data Influence the Estimates of Recently Arrived Immigrants in the American Community Survey

Elizabeth M. Grieco, Luke J. Larsen, and Howard Hogan

U.S. Census Bureau

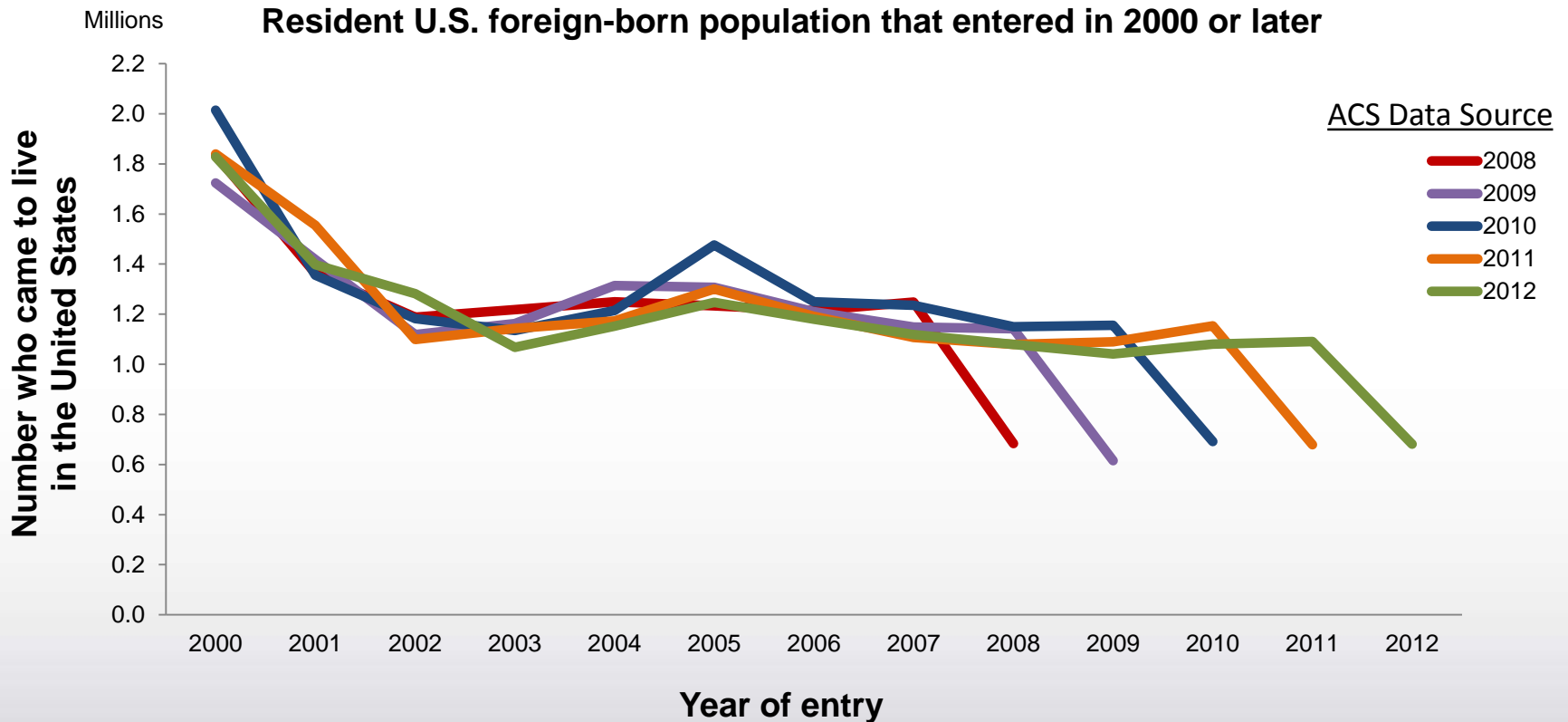## ACS Data Users Conference

Alexandria, Virginia

May 11-12, 2017

This presentation is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.

Immigrant's Year of Arrival

- The ACS asks: "When did this person come to live in the United States?"

- Asked of all foreign-born individuals (naturalized U.S. citizens or noncitizens).

- Responses given in specific, single years.

- Used to estimate the size and characteristics of "recent" arrivals to the United States.

Underestimation of "Most Recent" Arrivals?

Resident U.S. foreign-born population that entered in 2000 or later

Millions

Number who came to live in the United States

ACS Data Source
- 2008
- 2009
- 2010
- 2011
- 2012

Year of entry

Foreign-Born Population Who Reported Coming to Live in the United States in 2000 or Later by Year of Entry and 1-Year Data File: 2008 to 2012.
Source: U.S. Census Bureau, American Community Survey, 2008 to 2012 1-year files, internal data (weighted).

Some Possible Explanations

- Reluctance to participation in government surveys among new, unauthorized migrants (Myers, 2004).

- Circular migrants may be confused by meaning of "come to live" (Myers, 2004; Redstone and Massey, 2004).

- Coverage error in the ACS (Van Hook and Bachmeier, 2013).

An Alternate Proposal

The authors argue that the month-to-month data collection method used in the ACS program is the **primary** source of the apparent recent-arrival underestimates that can be observed in the data.

Structure of Presentation

- Background
  Features of ACS data and multi-year files.

- Methodology
  How relative timing of arrival and interview events in the ACS affects probabilities of inclusion and leads to underestimates of newly-arrived immigrants.

- Demonstration
  Observations of underestimation patterns evident in single- and multi-year ACS data files.

About the Data

Sources	Populations of Interest	Key Variables
• 1-yr ACS internal microdata • (2008-2012) • 3-yr ACS internal microdata • (2006-2012) • 5-yr ACS internal microdata • (2005-2012)	• Resident U.S. population • Foreign-born population • "Newly-arrived" foreign-born population	• Year of entry (YOE) • Year of survey (SY) • Month of survey/interview (IMO*)

All estimates were derived from population-controlled person weights (PWGT).

*Available via internal data only

United States Census Bureau
U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

7

ACS Data Collection

- The annual sample of housing units is divided into twelve segments (monthly panels).

- Each panel has three stages of data collection:
  - Month 1: Mail and internet modes
  - Month 2: All Month 1 nonresponse to telephone (CATI)
  - Month 3: Sample of Month 2 nonresponse to personal visits (CAPI)

- Panel month and interview month are not always the same.

Interpreting Period Estimates

- The multi-year estimates are NOT midpoints of the period or averages of the single-year estimates.

- Rather, they represent the average value of a population over the full period of the survey.
  - 2012 ACS 1-yr file: represents Jan 1, 2012 to Dec 31, 2012.
  - 2010-2012 ACS 3-yr file: represents Jan 1, 2010 to Dec 31, 2012.
  - 2008-2012 ACS 5-yr file: represents Jan 1, 2008 to Dec 31, 2012.

Timing of Immigrant's Arrival Relative to Interview

Two groups of foreign-born members of the ACS population sample in a given survey year:

- *Prior arrivals* – arrived prior to survey year
- *Recent arrivals* – arrived during survey year

For recent arrivals, the eligibility of being in the ACS sample is conditional upon the month of arrival relative to the month of the interview.

Immigrant's Probability of Inclusion (*p*) in ACS Sample

Given a housing unit selected for ACS sample:

- $p_{(Prior)}$ is generally constant (excluding mobility and mortality).

- $p_{(Recent)}$ varies with timing of interview, since arrival event can take place either before or after data collection has occurred.

- Simply put: $\qquad p_{(Prior)} \approx 1 \quad ; \quad 0 < p_{(Recent)} < 1$

- These are <u>not</u> sample selection probabilities; the housing units have already been selected for sample.

Windows of Opportunity

Scenario 1: A housing unit selected for sample in April Monthly Panel, and an immigrant moves into the housing unit at some point during the survey year.

Moves in <u>prior</u> to data collection.	Moves in <u>after</u> data collection.
Probability of interviewing recently arrived immigrant in household: approximately 4/12	Probability of <u>not</u> interviewing recently arrived immigrant in household: approximately 8/12

JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC

Windows of Opportunity

Scenario 2: A housing unit selected for sample in September Monthly Panel, and an immigrant moves into the housing unit at some point during the survey year.

Moves in <u>prior</u> to data collection.

Moves in <u>after</u> data collection.

Probability of interviewing recently arrived immigrant in household: approximately 9/12	Probability of <u>not</u> interviewing recently arrived immigrant in household: approximately 3/12

JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC

Formulation (Basic)

Let A be the event wherein an immigrant joins a sampled household prior to its interview. Then, $\Pr(A|m) = p_{(Recent)} \approx m/12$ for $m = 1, 2, ..., 12$.

It follows that:

$$\Pr(A) = \sum_{i=1}^{12} \Pr(A|m=i) * \Pr(m=i) \approx \frac{1}{12}\sum_{i=1}^{12} p_{(Recent)}$$

$$= \frac{1}{12}\sum_{i=1}^{12} \frac{i}{12} = \frac{78}{144} \approx 0.5417$$

Formulation (Refined)

When adjusting for partial-month values:

- $p_{(Recent)} \approx \dfrac{m-0.5}{12}$

- $\Pr(A) \approx \dfrac{1}{12}\sum_{i=1}^{12}\dfrac{i-0.5}{12} = 0.5$

So, recent immigrants whose housing units are part of the ACS sample in a given survey year have about a 50-50 chance to be counted among the sample under the monthly data collection scheme.

Underestimate of Recent Arrival Immigrant Counts

Using the $\Pr(A) = 0.5$ finding, it can be shown that the estimated count of survey-year immigrant arrivals in the ACS has an expected value that is downward biased by about 50 percent.

- Plausible value for single-year ACS data files under interpretation that the estimate represents contribution of person-years toward that value.

- However, this perspective does not hold when applied to multi-year ACS data files.

Compounding Bias in ACS Multi-Year Files

- The individual survey years that make up a multi-year file contribute data to the aggregate in roughly equal measure (e.g., ACS 2010, 2011, and 2012 each contribute about the same to the 2010-2012 3-year file).

- Problem occurs when estimating counts for most recent years of entry. For example, SY 2010 does not have YOE values of 2011, but it still contributes equally toward the YOE=2011 estimate.

Recent Arrival Estimates with ACS 3-year File

Data year	Year of Arrival Within Data Year					
2010	…	2007	2008	2009	2010	
			50%	**0%**	**0%**	
2011	…	2007	2008	2009	2010	2011
			100%	**50%**	**0%**	
2012	…	2007	2008	2009	2010	2011	2012
	B		**100%**	**100%**	A	**50%**

2012 1-year file: 2012 arrivals: 100 - **50** = 50% underestimation of expected arrivals

2010-2012 3-year file:
2010 arrivals: 100 - ((**50**/3) + (**100**/3) + (**100**/3)) = 17% underestimate of expected arrivals
2011 arrivals: 100 - ((**0**/3) + (**50**/3) + (**100**/3)) = 50% underestimate of expected arrivals
2012 arrivals: 100 - ((**0**/3) + (**0**/3) + (**50**/3)) = 83% underestimate of expected arrivals

United States Census Bureau
U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

18

Extension to ACS 5-Year File

Data year	Year of Arrival Within Data Year								
2008	…	2004	2005	2006	2007	2008			
						50%	0%	0%	0%	0%
2009	…	2004	2005	2006	2007	2008	2009		
						100%	50%	0%	0%	0%
2010	…	2004	2005	2006	2007	2008	2009	2010	
						100%	100%	50%	0%	0%
2011	…	2004	2005	2006	2007	2008	2009	2010	2011
						100%	100%	100%	50%	0%
2012	…	2004	2005	2006	2007	2008	2009	2010	2011	2012
					C	100%	100%	B 100%	100%	A 50%

2008-2012 5-year file:

2008 arrivals: 100 - ((**50**/5) + (**100**/5) + (**100**/5) + (**100**/5) + (**100**/5)) = 10% underestimation

2009 arrivals: 100 - ((**0**/5) + (**50**/5) + (**100**/5) + (**100**/5) + (**100**/5)) = 30% underestimation

2010 arrivals: 100 - ((**0**/5) + (**0**/5) + (**50**/5) + (**100**/5) + (**100**/5)) = 50% underestimation

2011 arrivals: 100 - ((**0**/5) + (**0**/5) + (**0**/5) + (**50**/5) + (**100**/5)) = 70% underestimation

2012 arrivals: 100 - ((**0**/5) + (**0**/5) + (**0**/5) + (**0**/5) + (**50**/5)) = 90% underestimation

# Recent Arrivals by Month of Interview (ACS 2012)

Thousands
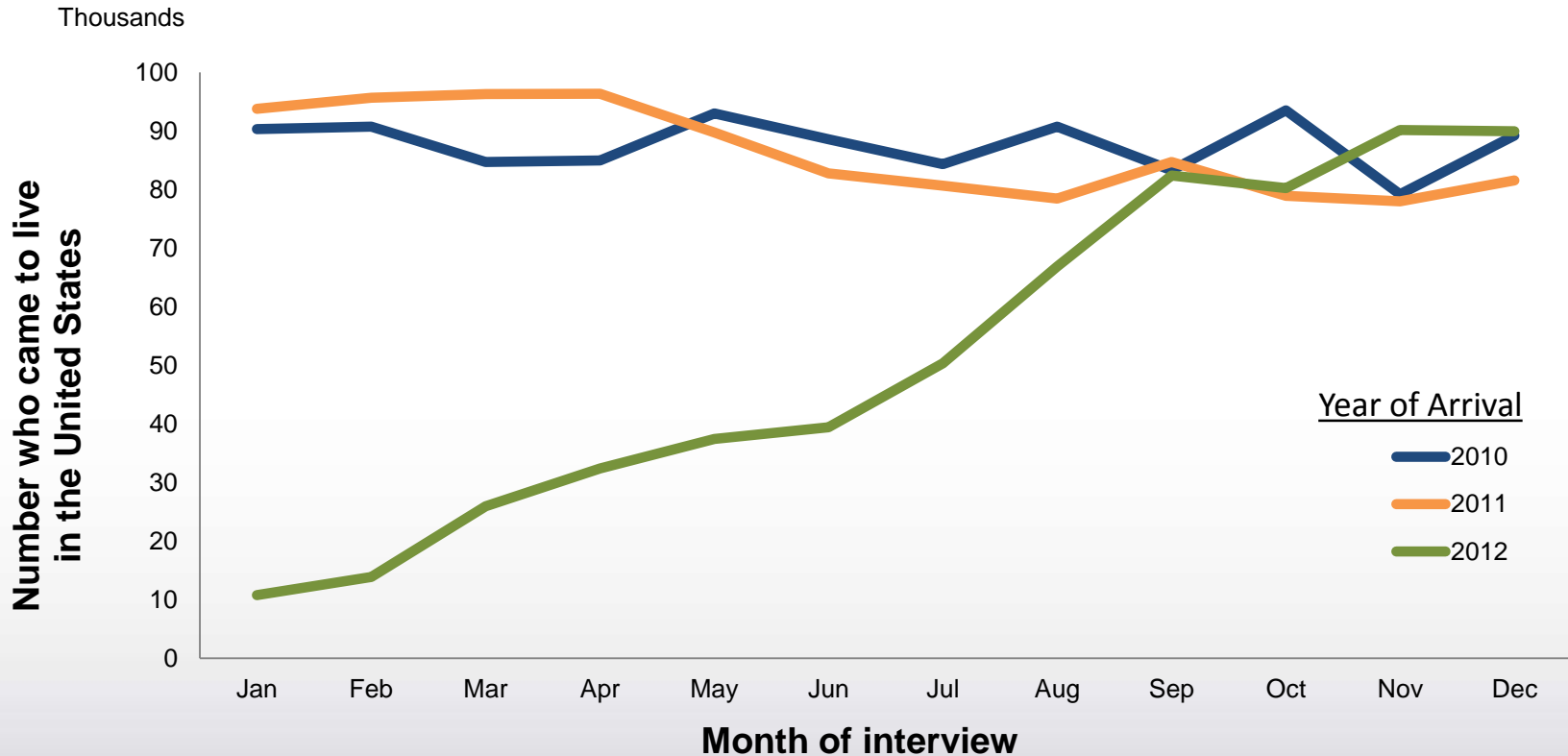
Reported Year of Entry (2010-2012) of the Foreign-Born Population Surveyed in 2012 by Month of Interview.
Source: U.S. Census Bureau, American Community Survey 2012 1-year file, internal data (weighted).

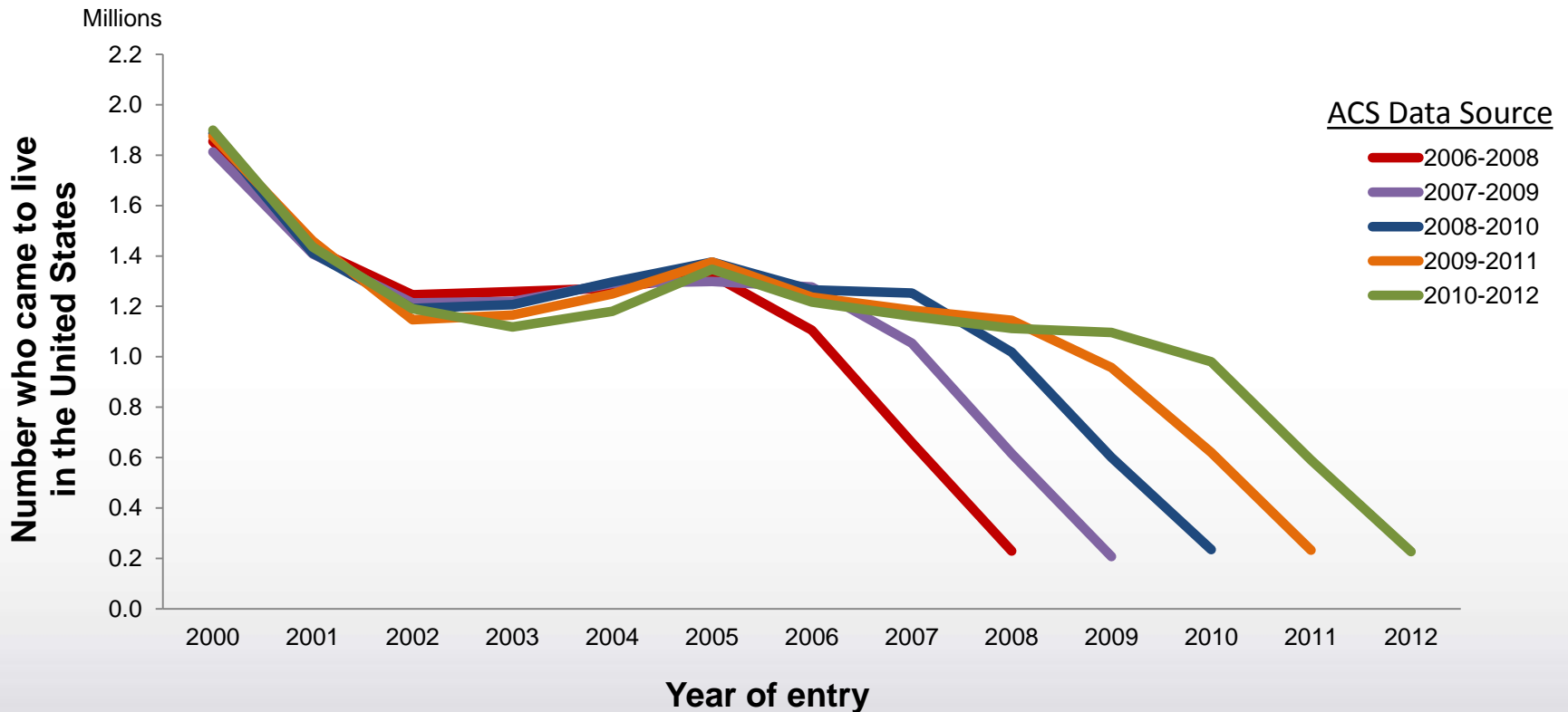# Year of Entry Among Single- and Multi-Year Files (ACS 2008-2012)
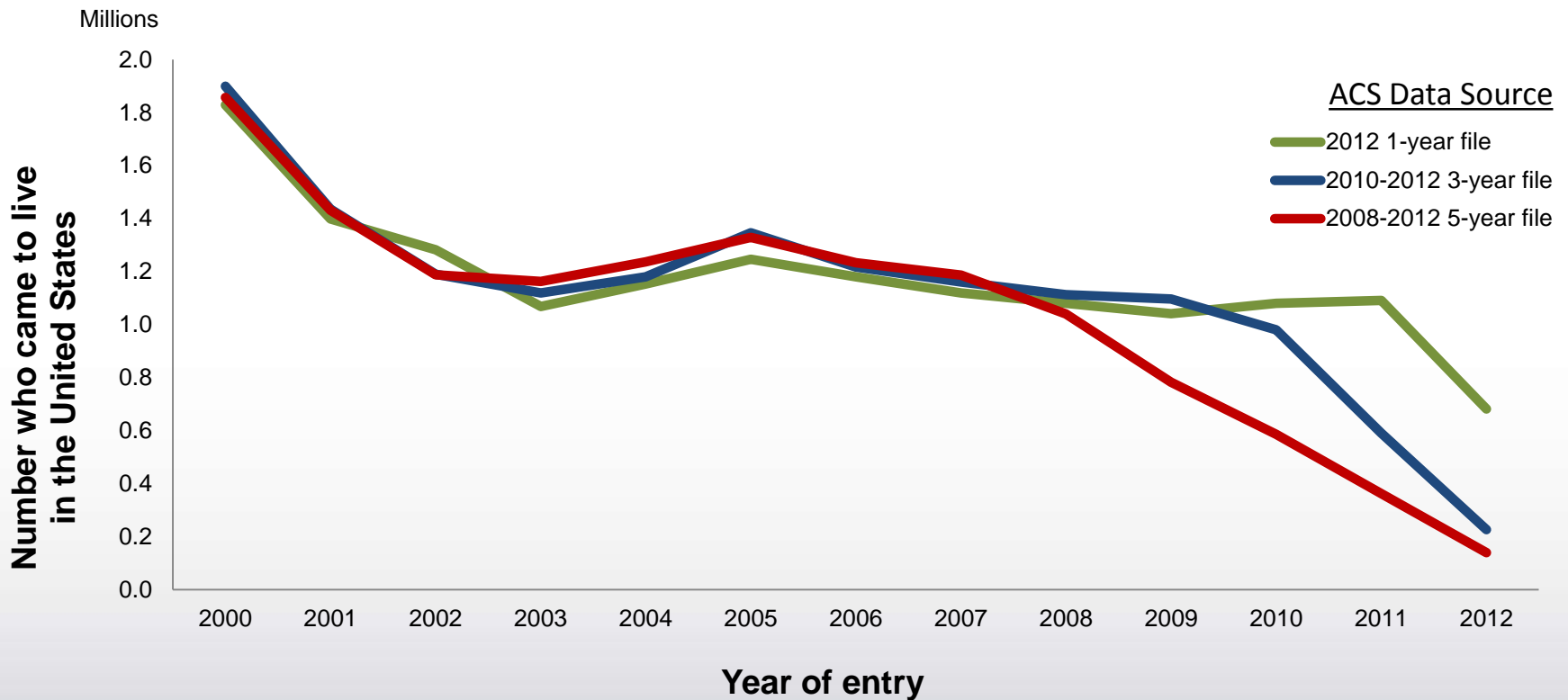
Foreign-Born Population Who Reported Coming to Live in the United States in 2000 or Later by Year of Entry and Data File: 2012, 2010-2012, and 2008-2012
Source: U.S. Census Bureau, American Community Survey, 2012 1-year, 2010-2012 3-year, and 2008-2012 5-year files, internal data (weighted).

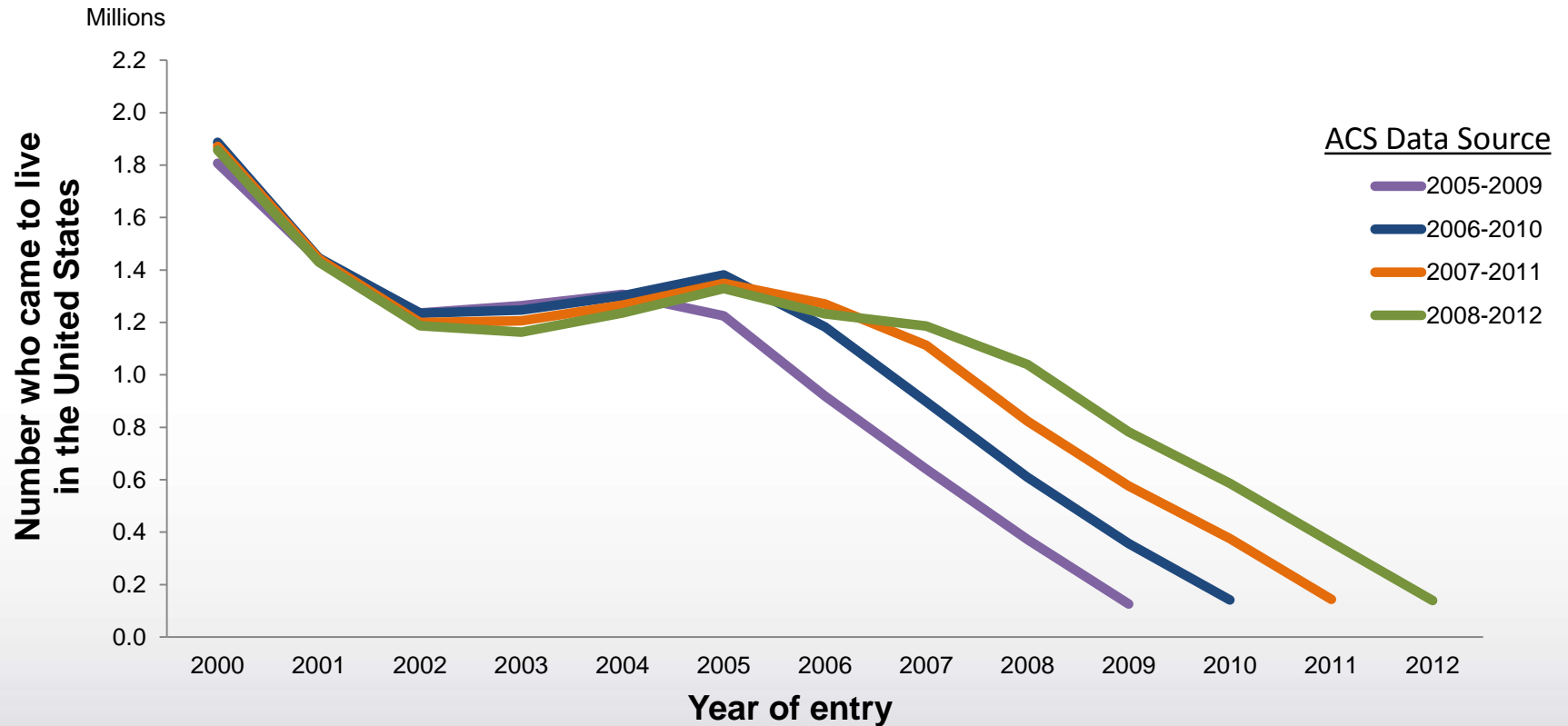# Year of Entry Among 1-Year Files (ACS 2008-2012)
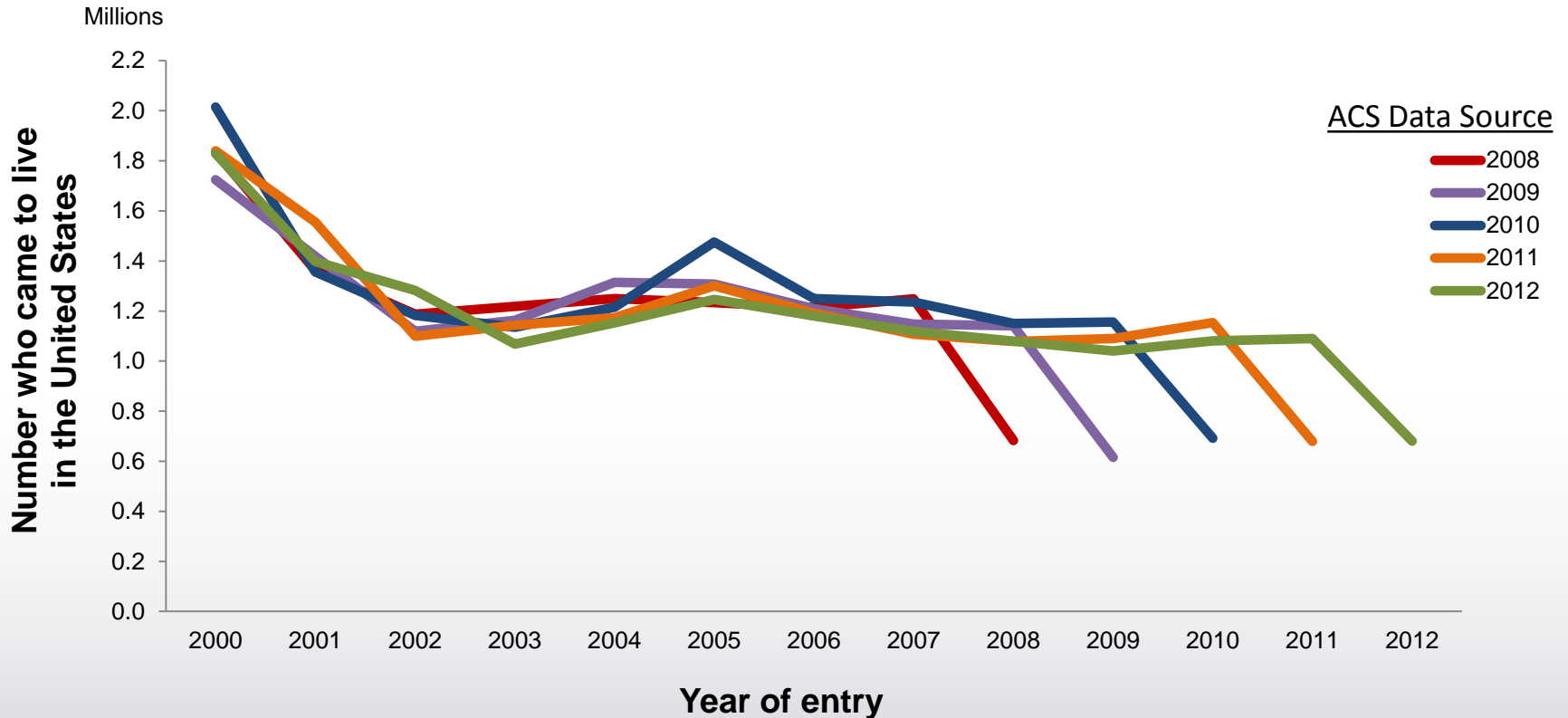
Foreign-Born Population Who Reported Coming to Live in the United States in 2000 or Later by Year of Entry and 1-Year Data File: 2008 to 2012.
Source: U.S. Census Bureau, American Community Survey, 2008 to 2012 1-year files, internal data (weighted).

Year of Entry Among 3-Year Files (ACS 2006-2012)

Foreign-Born Population Who Reported Coming to Live in the United States in 2000 or Later by Year of Entry and 3-Year Data File: 2008 to 2012.
Source: U.S. Census Bureau, American Community Survey, 2008 to 2012 3-year files, internal data (weighted).

United States Census Bureau
U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

23

Year of Entry Among 5-Year Files (ACS 2005-2012)

Foreign-Born Population Who Reported Coming to Live in the United States in 2000 or Later by Year of Entry and 5-Year Data File: 2009 to 2012.
Source: U.S. Census Bureau, American Community Survey, 2009 to 2012 5-year files, internal data (weighted).

Summary

- Year-long ACS data collection methodology (i.e., monthly panels) results in underestimation of newly arrived immigrants.

- The method of creating multi-year datasets compounds the bias for recent-arrival estimates.

- Researchers are urged to use caution when interpreting year-of-entry data, particularly for immigrants that arrived within the last five years.

Questions?

elizabeth.m.grieco@census.gov

luke.j.larsen@census.gov

howard.r.hogan@census.gov

United States **Census** Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov